



**HAL**  
open science

## Operationalizing variables

Amanda Edmonds, Aarnes Gudmestad

► **To cite this version:**

Amanda Edmonds, Aarnes Gudmestad. Operationalizing variables. Aarnes Gudmestad; Amanda Edmonds. Critical reflections on data in second language acquisition, John Benjamins Publishing Company, pp.125-148, 2018, 9789027201423. <10.1075/llt.51.06edm>. <hal-03051345>

**HAL Id: hal-03051345**

**<https://hal.science/hal-03051345v1>**

Submitted on 30 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

## **Operationalizing variables: The case of future-time expression in L2 French**

Amanda Edmonds

Université Paul-Valéry Montpellier 3

Aarnes Gudmestad

Virginia Polytechnic Institute and State University

### **Abstract**

In the field of second language acquisition (SLA), the importance of and the challenges associated with data-coding decisions often go unaddressed. And yet the operationalization of variables ultimately determines our capacity to both enter into dialogue with previous research and to address new and innovative research questions. In the current chapter, we reflect on data operationalization within the field of SLA, offering two concrete examples from investigations into the expression of future time in second-language French. We demonstrate how two important independent variables – temporal distance and adverbial specification – have either been defined in numerous ways or underspecified in the literature. We then provide concrete illustrations of the impact of data-coding decisions by presenting reanalyses of these two variables using data from previously published research. We conclude with a discussion of the implications that these critical assessments of data coding have for knowledge of SLA.

### **Introduction**

In the field of second language acquisition (SLA), we are often faced with the challenge of how to go about coding the data that are at the heart of any empirical approach to linguistics. It is crucial that data-coding decisions allow researchers to address new hypotheses and research questions, while at the same time allowing them to compare and

contrast with past research findings, thereby building the knowledge base within the field (Révész, 2012). This tension between new and sometimes innovative operationalization decisions and the need for generalization (and, thus, points of comparison with previous research) tends to go unaddressed in much research published in SLA. By neglecting to thoroughly attend to this crucial aspect within SLA research method head on, we inevitably head towards the curious situation of using the same (or similar) terms to talk about variables that have been operationalized in quite different manners (for examples in the field of phraseology, see Myles & Cordier, 2017; in the field of corrective feedback, see Ranta & Lyster, 2007). This state of affairs leads to claims of knowledge building that may be unfounded or overstated.

The current chapter reflects on the issue of data operationalization within SLA, using the specific example of future-time expression in second-language (L2) French as an illustration. We begin with a general reflection on the challenges associated with data coding, the potential for lack of consistency, and the dangers posed to the ability to generalize and draw valid conclusions. We support these observations with examples of data-coding decisions in variationist studies on future-time expression in native-speaker (NS) French (Blondeau, 2006; Blondeau & Labeau, 2016; Comeau, 2011; Grimm, 2010, 2015, 2016; Grimm & Nadasdi, 2011; Gudmestad, Edmonds, Donaldson, & Carmichael, in press; King & Nadasdi, 2003; Poplack & Dion, 2009; Poplack & Turpin, 1999; Roberts, 2012, 2016; Villeneuve & Comeau, 2016) and in L2 French (Blondeau, Dion, & Ziliak, 2014; Edmonds & Gudmestad, 2015; Gudmestad & Edmonds, 2016; Nadasdi, Mougeon, & Rehner, 2003). In the second portion of the chapter, we illustrate the significance of data-coding decisions by proposing two reanalyses of data from published studies. We conclude the chapter with a call for more explicitness in reporting and for clear justifications when linguists depart from

previously established data-coding practices, arguing that both are necessary in order to move forward SLA research in a meaningful way.

### **Challenges associated with data coding: The example of future-time expression in French**

In a discussion of frequent shortcomings in SLA method sections, Mackey (2012, p. 26) remarked that “coding systems vary widely and are not always represented in sufficient detail.” Concerning first the wide variety of data-coding practices in the field of SLA, Mackey and Gass (2005) suggest that this can lead to lack of comparability across different studies. For this reason, they advise researchers to adopt existing data-coding schemas, unless changes in those schemas are clearly motivated. The second observation in the previously cited quote highlights the fact that descriptions of data coding are not always sufficiently precise. While this is sometimes the result of constraints over which researchers have no control (e.g., length limits placed on publications), the results of this lack of precision can be serious and weaken the ability of the scientific community to judge the importance of the study in question and to make relevant comparisons across studies. Lack of clarity may exist with respect to what is being studied (i.e., the dependent variable), as well as how the phenomenon under study was analyzed (e.g., independent variables). Thus, the variety in practices and the fact that sufficient details are not always provided together have the potential to undermine the ability to interpret the findings presented, as well as the potential for replication (Mackey, 2012; Marsden, Mackey, & Plonsky, 2016; Polio & Gass, 1997; Révész, 2012). In other words, the fuzziness that exists with respect to data coding can impede attempts at building knowledge that is generalizable and verifiable within the field. Moreover, while any investigation into linguistic phenomena may suffer from data-coding problems, the coding of interlanguage

poses particular challenges, given its “unstable” and “transitional” nature (Corder, 1981, p. 18).

Although we expect that the preceding comments will not be particularly controversial, the actual impact of data-coding decisions is rarely given center stage in the literature (for a recent exception, see Myles & Cordier, 2017). In the current chapter, we provide two concrete illustrations of how coding decisions may influence results and, ultimately, constrain conclusions drawn with respect to SLA. To do so, we have chosen the domain of future-time expression in L2 French, an area that we have researched within variationist (Geeslin & Long, 2014) and concept-oriented approaches (von Stutterheim & Klein, 1987) to SLA (see the sections *Future-time expression* and *Reanalyzing adverbial specification, respectively*, for definitions of these two approaches). When we began examining future-time expression, we were struck by the high level of diversity with respect to data-coding practices and reporting, diversity that characterized both the identification of the dependent variable (i.e., what counts as a future-time context) and the operationalization of numerous independent variables. This variety is present in studies of both first-language (L1) and L2 French and complicates our ability to engage in dialogue with previously published research. Faced with multiple ways of operationalizing certain variables, for which coding practices were not always sufficiently explicit, we found ourselves struggling with these decisions. We have published reflections on the coding of the dependent variable in a previous article (see Edmonds, Gudmestad, & Donaldson, 2017). In the current chapter, we use the example of two of the most frequently studied independent variables – namely, temporal distance (also sometimes referred to as temporal reference) and the presence or absence of a lexical temporal indicator (LTI, often referred to as adverbial specification) – in order to highlight these challenges and, then, to examine how different decisions lead to different findings. In the remainder of this section, we offer a concise presentation of variable

future-time reference in French, followed by an in-depth discussion of how temporal distance and LTIs are thought to influence this case of variation. In particular, we highlight differences in how these two factors have been operationalized in previous variationist research. Finally, we present the rationale for the current chapter.

### *Future-time expression in French*

Future-time reference may be expressed in French using a variety of verb forms. Whereas Edmonds et al. (2017), using a concept-oriented approach, identify no fewer than 13 such forms in the NS portion of a corpus of informal conversations, most previous studies have focused on two or three of the most frequent verb forms found in future-time contexts. These include the periphrastic future (PF: 1a), the inflectional future (IF: 1b), and the futurate present (1c).

(1) a. *je pense qu'il va le faire* (NS5, t249)<sup>1</sup>

'I think that he is going to do it'

b. *mais tu feras ça samedi* (NS5, t225)

'but you will do it Saturday'

c. *qu'est-ce que tu fais demain après-midi* (NS5, t269)

'what are you doing tomorrow afternoon'

In each of these three examples, the same NS is making reference to future time with the verb *faire* 'to make/do', although three different verb forms are used. This particular case of alternation has received much attention, particularly from researchers interested in examining morphosyntactic variation within a variationist perspective. A variationist approach to future-time expression recognizes that certain instances in which reference to future time is made can be expressed using more than one form (called *variants*). Those instances in which variation is possible make up the *envelope of variation*, and variationist studies generally

---

<sup>1</sup> Examples come from the corpus used in Edmonds et al. (2017). 'NS5 refers to the fifth native speaker in the corpus and the token number 249 in the dataset.

employ quantitative methods (especially regression models) with the aim of identifying the internal and external factors that influence frequency of use of a given variant within that envelope (Tagliamonte, 2012).

While variationist linguistics began in the 20<sup>th</sup> century (Labov, 1966), the interest in variation in the future-time sector in French can be traced back hundreds of years. As shown by Poplack and Dion (2009), grammarians writing between 1530 and 1999 addressed the variation in future-time expression, most often attempting to explain it away by evoking factors that constrain which variant will be used. Poplack and Dion write that “[m]ost grammarians, prescriptive and descriptive, espouse the idea that the variants are selected according to the way the future eventuality is envisioned, and/or the semantic or pragmatic import to be conveyed” (p. 558). In the sentence that follows, Poplack and Dion note that “[i]n spontaneous speech, however, [the variants] are rarely used in accordance with the values proposed.” This observation is the starting point for variationist analyses that attempt to understand which linguistic and extralinguistic factors may influence the verbal variant used by a given speaker. Although the operationalization of extralinguistic factors is also worthy of attention, in the current chapter, we concentrate only on linguistic factors. With respect to future-time reference, linguistic factors that are commonly included as independent variables are temporal distance, presence or absence of an LTI, sentential polarity, the certainty that the future event will occur, grammatical person, and, to a lesser extent, contingency on an *if* clause, and the presence of *quand* ‘when’. Whereas the operationalization of certain independent variables is relatively uncontroversial (e.g., the grammatical person of the subject is generally clear), other variables are more open to interpretation. This is particularly the case with the first two variables mentioned, namely temporal distance and LTIs. In what follows, we review how temporal distance and LTIs have

been operationalized in the literature on L1 and L2 future-time reference in French, generally limiting the discussion to research on oral data conducted within a variationist perspective.

#### *Future-time reference and temporal distance*

The distance between time of speaking and the expected moment at which the future event (be it an action or a state) will be realized has long been thought to influence which verb form is used to make reference to future time. This is particularly clear with the PF, which is also called the *futur proche* or *futur prochain* ‘close future’, revealing the presumed connection between this form and events that are set to occur in the near future. In their review of 163 grammars, Poplack and Dion (2009) found that “the proximity reading for PF enjoys the highest rate of agreement (59%) among grammarians” (p. 574). However, when the presumed connection between PF use and temporal distance has been examined in actual language use, findings diverge along geographic lines. Comeau and Villeneuve (2016, p. 234) go so far as to speak of a “temporal distance-polarity divide” when it comes to linguistic constraints on future-time reference. As they point out, it has generally been found that future-time reference in varieties of French spoken in Québec, Ontario, and Western Canada are most strongly conditioned by sentential polarity (e.g., Blondeau, 2006; Grimm, 2015<sup>2</sup>; Grimm & Nadasdi, 2011; Poplack & Turpin, 1999; Poplack & Dion, 2009), whereas the expression of future time in Acadian varieties of French has been found to be constrained by temporal distance (Comeau, 2011; King & Nadasdi, 2003). As for French spoken in France, the results for these two variables are currently mixed: Roberts (2012) reports that sentential polarity – but not temporal distance – is operative, Villeneuve and Comeau (2016) find that only temporal distance significantly influences future-time expression, and Gudmestad et al. (in press) find that both variables are significant. As for the role played by temporal distance in

---

<sup>2</sup> In his dissertation, Grimm looked at a corpus from 1978 and one from 2005, as well as data from several different communities in Ontario. Most analyses showed sentential polarity (and not temporal distance) to be operative. However, certain types of speakers (i.e., restricted speakers in the 1978 corpus) and certain communities (i.e., Cornwall, North Bay, and Pembroke) were found to be sensitive to both sentential negation and temporal distance.

future-time expression L2 French, we know of two variationist studies that have looked at oral data. Whereas Blondeau et al. (2014) found that the PF was significantly favored in the expression of proximal future events in the speech of Anglo-Montrealers, this factor had no significant effect in Nadasdi et al.'s (2003) analysis of speech produced by students living in Ontario and attending a French immersion program.

It bears noting that this body of evidence regarding temporal distance is based on at least seven different ways of dividing up the temporal distance spectrum, from the binary *proximal* versus *distal* opposition seen in (2a), to the six temporal categories coded by Roberts (2013 and 2016, see [2g]). Clearly, King and Nadasdi (2003, p. 333) were correct when they wrote that “the exact definition of what constitutes ‘near’ is difficult to pin down.”

- (2) a. Proximal (< 24 hours) versus distal (> 24 hours) (Blondeau, 2006, p. 86; Blondeau et al., 2014, p. 679; Blondeau & Labeau, 2016, p. 252; Grimm, 2015, p. 236; Grimm, 2016, p. 7; Nadasdi et al., 2003, p. 207; Poplack & Dion, 2009, p. 571; Poplack & Turpin, 1999, p. 150; Roberts, 2012, p. 99; Villeneuve & Comeau, 2016, p. 325, fn. 9)
- b. < 1 hour, > week, indeterminate, continuous (Grimm & Nadasdi, 2011, pp. 178-179)
- c. < 1 hour, < 24 hours, < week, > week, continuous (King & Nadasdi, 2003, p. 328)
- d. < 1 hour, < 24 hours, < week, > week, unspecified (Grimm, 2010, p. 86)
- e. < 24 hours, < week, < month, > month, ambiguous (Gudmestad et al., in press)
- f. < 1 hour, < 24 hours, < week, < year, > year, indeterminate (Comeau, 2011, pp. 218-219)
- g. < 24 hours, < week, < year, > year, continuous, indeterminate (Roberts, 2013, p. 142; Roberts, 2016, p. 293)

There is a clear tendency in this literature to opt for a two-way distinction between *proximal* and *distal*. However, this trend must be nuanced by at least three observations concerning the

studies mentioned in (2a). First, there is some ambiguity around the coding of future events for which temporal distance cannot be determined. Whereas many linguists opt for an explicit “indeterminate” or “ambiguous” category, others appear to have grouped indeterminate examples with distal ones (Blondeau & Labeau, 2016, p. 253; see also King & Nadasdi, 2003, for a similar coding decision). Other researchers simply have not addressed the question of indeterminate future-time reference in their publications. Second, certain linguists (such as Nadasdi et al., 2003) use the terms *proximal* and *distal* but do not define what is meant by these terms. Although we presume, given the previous studies they cite, that Nadasdi et al. have adopted the within a day / beyond a day distinction, the fact that this is not made explicit is problematic, especially when we consider that other authors working on future-time reference in French may use these terms with different definitions. This was the case, for example, in Moses’ (2002) concept-oriented analysis in which *proximal* referred to events occurring within the week following speech time. Third, Poplack and Turpin (1999), Roberts (2012), Grimm (2015, 2016), and Villeneuve and Comeau (2016) originally opted for a finer-grained coding of temporal distance, only later collapsing to a binary distinction either because there was no evidence that the finer coding provided additional insight (Poplack & Turpin) or because the distribution of the data did not allow the researchers to retain the more detailed coding (Roberts; Villeneuve & Comeau).<sup>3</sup> Thus, if a majority of studies has ultimately opted for the binary opposition, this distinction is not always the first coding choice preferred by scholars.

The tendency to start with a multi-categorical variable before reducing to a binary or ternary opposition is relatively common in this literature, where collapsing categories appears most often to be motivated by the need to avoid small cells in the regression analyses that characterize variationist linguistics. In such cases, and space permitting, it would be useful for

---

<sup>3</sup> Grimm simply states that “it was not necessary to scrutinize the data to this level of detail” (2015, p. 236).

authors to also report the tendencies found with the original coding, as was done by Villeneuve and Comeau (2016) with respect to temporal distance. After having presented their multivariate analysis with a binary temporal distance variable, these authors returned to their original coding for a follow-up analysis. In their own words, they recognize that

the original finer breakdown reveals a finding that is unfortunately masked by the broader binary coding necessary for the multivariate analysis: the high frequency of the periphrastic future in the most immediate contexts. [...] These results show that the PF is used at an extremely high rate (93.3%) with events anticipated to occur within a minute, which suggests that the PF is marking imminence, and appears to confirm early grammarians' descriptions: the periphrastic future does mark *le futur proche*. (p. 329)

In discussing this finding, the researchers highlight the difficulty in comparing results across studies: “In the other varieties [of French] which display a strong temporal distance effect, it is impossible to establish whether or not the PF is as highly favoured in imminent contexts (i.e., within the minute) due to the fact that the authors did not adopt as fine a temporal breakdown” (p. 330). In other words, comparison of results concerning temporal distance is compromised by the wide variety of coding schemas adopted in this literature. Moreover, the justification for the different ways of coding temporal distance presented in (2a-g) has, to the best of our knowledge, not been addressed.

#### *Future-time reference and LTIs*

Temporal adverbials (also referred to as LTIs in this chapter) that have the possibility of being future-looking, such as *tout de suite* ‘right away’, *demain* ‘tomorrow’, and *un jour* ‘one day’, are one means of establishing reference to future time. The importance of such LTIs in modern-day French has been extensively discussed with respect to future-time reference, and especially as concerns the futurate present (e.g., Grimm, 2016; Le Goffic &

Lab, 2001). The futurate present does not carry explicit future morphology, and it has been suggested that future-time reference must be established by other means, such as temporal adverbials (see Blondeau, 2006; Moses, 2002). In recent variationist accounts of future-time reference, datasets have been coded for the presence versus absence of temporal adverbials, with most linguists making a three-way distinction between absence of an adverbial, presence of a specific adverbial (e.g., *demain* ‘tomorrow’), and presence of a non-specific adverbial (e.g., *bientôt* ‘soon’). Research that has conducted multivariate analyses for only the PF and the IF in future-time contexts has most often found no significant effect for adverbial specification (Blondeau, 2006; Grimm & Nadasdi, 2011; King & Nadasdi, 2003; Roberts, 2012, 2016; Villeneuve & Comeau, 2016). Poplack and Dion (2009), Comeau (2011), and Blondeau and Labeau (2016) constitute exceptions, and in each of these studies, the IF was found to be favored in contexts containing an LTI.

At least five studies have used regression models to examine the futurate present, in addition to the PF and IF, in the expression of future time by NSs or non-native speakers (NNSs) of French: Poplack and Turpin (1999) using the Ottawa-Hull corpus, Grimm (2016)<sup>4</sup> for the community of Hawkesbury (Ontario), Gudmestad et al. (in press) for Hexagonal French, Nadasdi et al. (2003) for French spoken by Canadian school children in an immersion program, and Blondeau et al. (2014) for the French of Anglo-Montrealers. The results from these five studies are summarized in Table 1, which illustrates that the present has consistently been favored in the presence of a temporal adverbial, for both NSs and NNSs.

---

<sup>4</sup> Grimm (2015) also analyzed the present-for-future in several different datasets. Because of space restrictions, we choose here to present only the results reported in his 2016 article.

Table 1. *Results from variationist studies having examined the PF, IF, and present in oral production*

Study	Population	PF	IF	Present
Poplack & Turpin (1999)	NSs (Ottawa)	Favored when LTI is absent	Favored when a non-specific LTI is present	Favored when a specific or non-specific adverbial is present
Grimm (2016)	NSs (Ontario)	Favored when LTI is absent	Non-significant	Favored when a specific adverbial is present
Gudmestad et al. (in press) <sup>5</sup>	NSs (France)	Presence of specific LTI: IF < Present indicative (non-significant for non-specific LTI) Presence of specific or non-specific LTI: PF < Present indicative		
Nadasdi et al. (2003)	NNSs (Ontario)	Favored when LTI is absent	Non-significant	Favored when a specific or non-specific adverbial is present
Blondeau et al. (2014)	NNSs (Québec)	Favored when LTI is absent	Favored when a non-specific LTI is present	Favored when a non-specific LTI is present

<sup>5</sup> Results from this study look different as a multinomial regression was run. This type of analysis makes a three-way dependent variable possible, allowing us to examine the full dataset in a single model. For the other four studies, researchers ran three different models (one for each form of the dependent variable).

On its surface, the operationalization of this factor appears quite straightforward: A future-time context either is or is not accompanied by a temporal adverbial making reference to future time. However, the difficulty – and the ambiguity – lie precisely in what is meant by *accompanied*. As we began coding our own data, this ambiguity became apparent, and we questioned how close a temporal adverbial had to occur to a future-time context in order to count as *present*. Did the temporal adverbial specifying when a future event was to occur have to appear within the same clause, or could it occur within the same utterance, the same turn, or even within the same discourse context as the verb expressing the event in question? Unfortunately, previous studies that have provided details about the coding of this factor are rare: Poplack and Turpin (1999) and Poplack and Dion (2009) state that adverbial specification was coded at the level of syntax in their studies, whereas Grimm (2015, pp. 229-230) considered that a temporal adverbial was present if used in the same turn as the future-time context or in the immediately preceding question by the interviewer. In the remaining studies, data-coding practices have not been made explicit, although the examples cited show temporal adverbials occurring in the same clause as the future-time context. It is possible that coding in these studies actually went gone beyond the clause level, and that space constraints led authors to choose short examples. However, given that descriptions of coding did not clarify this possibility, we deduced that most previous studies had likely adopted a clause-level coding for LTI, as suggested by the examples provided. This is thus how we ultimately operationalized this factor for Edmonds et al. (2017) and Gudmestad et al. (in press).

Opting for a clause-level coding of LTI could be justified on practical grounds, but perhaps less so on theoretical ones. Generally speaking, data-coding schemas should be both reliable (i.e., allow for consistent categorization of data) and theoretically valid (Révész, 2012). A clause-level coding of LTI certainly allows for consistent categorization of the variable, constituting an excellent example of what Révész (2012, p. 213) refers to as “low-

inference categories,” that is coding categories “which require little judgment” on the part of the coder. However, the justification of clause-level coding on a more theoretical level is less clear. Indeed, restricting this factor to the level of the clause effectively implies that the presence of temporal adverbials within the larger discourse beyond the clause does not influence verb forms used to express future time. Although this may very well be the case in interview data, in looking at our own conversation data, we saw numerous instances of future-time reference being established via a temporal adverbial and being maintained across several turns, without repetition of LTIs. A short extract from an example that will be analyzed later in the paper illustrates this phenomenon (3):

- (3) NS1: *et le vingt-huit **il y a** une avant-première **y a** une pièce de théâtre du collègue Jeanne d’Albret*  
‘and the 28<sup>th</sup> **there is** premiere **there is** a play by the Jeanne d’Albret middle school’  
NNS1: *ah ouais **j’y vais** et Chloé **elle joue**?*  
‘ah yeah **I’m going** and **is Chloé performing**?’  
NS1: *ouais **Chloé y joue** ouais*  
‘yeah **Chloé’s performing** yeah’

Here, the NS establishes future-time reference with the LTI *le vingt-huit* (the 28<sup>th</sup>), and this temporal frame is maintained over the next several clauses and turns, by both the NS and the NNS. In light of such examples, we question whether a clause-level definition of LTI, which appears to correspond to current practice in this literature, actually reflects how temporal adverbials are used in conversation data in order to establish and maintain temporal reference. These reflections have led us to explore the possibility of offering a new operationalization of this variable (see Grimm, 2016, for a discussion of a similar issue in interview data and Gudmestad et al., in press, for a first attempt at addressing this issue with conversation data).

*Rationale*

As demonstrated throughout this section, two important variables thought to influence future-time reference in French – namely, temporal distance and adverbial specification – have been subject to a variety of data-coding decisions. Temporal distance has been sliced into anywhere from two to six categories of various temporal lengths, clearly complicating our ability to build knowledge across studies. The issue at hand with adverbial specification is somewhat different, as what is meant by *presence* and *absence* of an LTI has largely gone unaddressed. In other words, in these two variables, we find both an example of “coding systems [that] vary widely” and one of coding systems that “are not always represented in sufficient detail” (Mackey, 2012, p. 26). In what remains of this chapter, we offer two reanalyses of previously published data. First, we explore how the operationalization of temporal distance may change outcomes in terms of findings reported for an elicitation task (Edmonds & Gudmestad, 2015). We then return to conversation data that we analyzed in Edmonds et al. (2017) in order to offer an alternative approach to operationalizing adverbial specification. In both cases, we discuss the implications that these critical assessments of data coding have for knowledge of L2 acquisition.

### **Reanalyzing temporal distance**

As shown previously, the factor of temporal distance has been defined in numerous ways in studies looking at future-time reference in French, with the binary opposition between proximal (< 24 hours) and distal (> 24 hours) being the most frequently adopted. In this section, we explore the impact of two different operationalizations of this variable on a single dataset (Edmonds & Gudmestad, 2015). We will begin by presenting the study, in which temporal distance was originally divided into five categories. For the current chapter, we recoded the variable of temporal distance, collapsing it into a two-way distinction, before reanalyzing the dataset. We present these new findings and then end by comparing the results.

### *Presentation of original study*

For our first publications on future-time expression, we opted to use a data-elicitation task in order to examine the role that three linguistic factors may play in future-time expression: temporal distance, LTIs, and (un)certainty markers (see IRIS repository for access to the task). The written contextualized task that we created contained a series of 30 short contexts that together told a story (for the original, Spanish version of the task, see Gudmestad & Geeslin, 2013). Each context ended with an expression that introduced direct dialogue (e.g., “he said”, “she asked”), followed by three possible formulations for the dialogue. The formulations differed only in the form of the verb (PF, IF, or present), and the order in which the forms was presented was counter-balanced across items. As an example, item 24 is provided in (4):

(4) *Sarah demande à André ce qu’il compte faire après ses études. Comme Sarah, il finit dans deux ans. Il répond :*

- a. *Je cherche sans doute un travail en France.*                    \_\_\_\_ *Je préfère la phrase A.*  
b. *Je vais sans doute chercher un travail en France.*            \_\_\_\_ *Je préfère la phrase B.*  
c. *Je chercherai sans doute un travail en France.*                \_\_\_\_ *Je préfère la phrase C.*

Sarah asks André what he’s planning on doing after his studies. Like Sarah, he finishes in two years. He responds:

- a. I am without a doubt looking for work in France.            \_\_\_\_ I prefer sentence A.  
b. I am without a doubt going to look for work in France. \_\_\_\_ I prefer sentence B.  
c. I will without a doubt look for work in France.                \_\_\_\_ I prefer sentence C.

For each item, participants were asked to choose the formulation that they preferred. Across the 30 contexts, we manipulated temporal distance, LTIs, and (un)certainty markers. With respect to temporal distance, five different categories were examined: immediate, < 24 hours,

< week, < month, and > year. The temporal distance of the future event was made clear in each context, and the five categories were evenly distributed across the 30 items, resulting in six items per temporal distance category. This task was completed by 30 NSs of French and a total of 116 NNSs with 24 different L1s, all of whom were studying in France at time of testing (see Gudmestad & Edmonds, 2016, for additional analyses of this task). The NNSs were assigned to four levels of proficiency on the basis of a c-test (see Tremblay, 2011, for a discussion of c-tests as proficiency measures). This type of task presents learners with a passage in the target language from which a certain number of words has been deleted (50, in the case of the c-test used in the project in question). Learners were provided with half of the letters of the missing word and asked to complete all blanks.

In order to analyze the responses, a multinomial regression was conducted. The advantage of this type of analysis, when compared to binary regressions used in most variationist studies of future-time reference, is that a single model of the dataset is generated, thus taking into account the relationships between the three categories of the dependent variable (IF, PF, present). In Table 2, we present the temporal-distance results from Edmonds and Gudmestad (2015); for the full models, see the original study. In this type of model, both the dependent and independent variables have a base category against which the other categories are compared. In the model presented here, the base category for the dependent variable was the PF, with the results for the IF versus PF comparison presented on the left and the results for the present versus PF comparison presented on the right. As concerns the independent variable of temporal distance, < 24 hours was the reference point. For each comparison with this base, the odds of choosing the comparison category of the dependent variable can be higher, lower, equal, or not significant.<sup>6</sup> As an example, we see in the first

---

<sup>6</sup> When an independent variable is included in the regression model, this usually means that there is at least one significant effect for which the odds of choosing the comparison category of the dependent variable are lower or higher than the base case. In some cases, though, some categories of the variables are not shown to be significantly different from each other. Thus, when the independent variable is included in the model but a given

column that for Levels 1 and 3, the odds of choosing the IF over the PF are the same for an action set to occur in the immediate future (vs. one set to occur within 24 hours). However, for Levels 2 and 4 and for the NSs, the odds of choosing the IF over the PF are lower when the future event is expected to occur immediately, when compared to events that should occur within 24 hours.

Table 2. *Results for the temporal distance variable (Edmonds & Gudmestad, 2015)*

Group	IF vs. PF					Present vs. PF				
	Immed	< 24h	< week	< month	> year	Immed	< 24h	< week	< month	> year
Level 1	=	Base	=	=	>	=	Base	<	=	=
Level 2	<	Base	=	>	>	=	Base	<	=	=
Level 3	=	Base	=	>	>	=	Base	<	<	<
Level 4	<	Base	=	>	>	=	Base	=	<	<
NSs	<	Base	=	>	>	=	Base	<	=	<

These results led us to two conclusions. First, when examined within the context of the full model, these findings suggested that patterns of variation similar to those seen for NSs developed more slowly for these learners with respect to the present-for-future than for the IF, compared to the PF. More specifically, NNSs at the higher proficiency levels showed close approximation to NS patterns of variation in the IF versus PF comparison, whereas patterns identified when selection of the present was compared with that of the PF showed more variability (for all three independent variables examined). The second conclusion concerned the factor of temporal distance more specifically. In examining the IF versus PF comparison, we noted that the NSs dispreferred the IF in proximal contexts (i.e., immediate contexts), whereas this form was preferred in more distal ones (< month and > year). This pattern is also

---

category of that variable is not shown to be significantly different from another, the odds are found to be “equal.” In contrast, when an independent variable is not included in the regression model, this means that the factor does not condition the dependent variable and the variable is found to be “not significant.”

instantiated in the learner data, where we see non-linear development across levels. For example, the learners at the lowest level of proficiency (Level 1) prefer the IF to the PF in the most distal contexts (> year), with no differences seen for the other categories. Level 3 participants prefer the IF to the PF in the two most distal contexts (< month, > year) but show no significant preference between the two forms identified for the other categories. Finally, Levels 2 and 4 show the NS pattern, in which the IF is disfavored in the immediate contexts and favored in the two most distal contexts. These findings may mean that, with regard to the IF-PF comparison, learners first show sensitivity in the expression of the most distal events before developing sensitivity to IF versus PF use in proximal ones.

### *Reanalysis*

For our reanalysis, we examined the same dataset using the most common operationalization of temporal distance: proximal (< 24 hours) versus distal (> 24 hours). This means that each participant responded to 12 proximal contexts (6 immediate, 6 < 24 hours) and 18 distal ones (ranging from < week to > year). We then ran a second multinomial regression. The results for temporal distance analyzed as a binary variable are presented in Table 3 (the results for the two other variables analyzed in this project did not change from those found in Edmonds & Gudmestad, 2015).

Table 3. *Results for the temporal distance variable reanalyzed*

Group	IF vs. PF		present vs. PF	
	Proximal	Distal	Proximal	Distal
Level 1	Base	>	Base	<
Level 2	Base	>	Base	<
Level 3	Base	>	Base	<
Level 4	Base	>	Base	<
NSs	Base	>	Base	<

When reanalyzed as a binary variable, we find that all four groups of learners and the NSs preferred the IF (over the PF) in distal contexts, whereas in those same contexts, the present was dispreferred (compared with the PF). In other words, the NSs and all four learner groups show preference for the IF and dispreference for the present in distal contexts. Whereas this general tendency is also visible in the results provided in Table 2, the two analyses differ crucially with respect to the patterns of development. Indeed, in the reanalysis using a binary proximal versus distal opposition (Table 3), all groups of participants showed similar selection patterns. Thus, even the learners with the lowest levels of proficiency match the NS patterns, meaning that no conclusions concerning development can be drawn. The analysis presented in Table 2, on the other hand, shows that a finer division of the temporal distance spectrum reveals developmental patterns, particularly in the IF versus PF comparison: the IF is preferred first in the most distal contexts (see Level 1) before becoming dispreferred in the most proximal ones (see Levels 2 and 4). Both findings are of interest to the field of SLA: (a) learners appear to be able to approximate the macro-distinction between proximal and distal contexts in an elicitation task even at low levels of proficiency (Table 3) and (b) at a finer level of distinction, developmental patterns are evident (Table 2). These findings obviously derive directly from data-coding decisions, and those decisions reflect what a researcher judges to be most important: in this case, a high level of detail versus comparability with previous studies. Whereas our original study gave preference to the use of “as finely grained a measurement as possible” (Mackey & Gass, 2005, p. 230), thus revealing information about development, the reanalysis sought to facilitate comparisons with previous studies by adopting a similar coding scheme. In so doing, the reanalysis allowed us to more easily note that patterns from our dataset are more in line with Blondeau et al.’s (2014) results from oral interviews conducted with Anglo-Montrealers, in which they found that the PF was favored in proximal contexts, than those reported by Nadasdi et al. (2003), who found no significant

effect for temporal distance. Thus, the issue at stake is not which one of the two coding decisions is “better”, but rather to demonstrate empirically that these two examples lead to different conclusions about L2 acquisition.

### **Reanalyzing adverbial specification**

Most variationist analyses into future-time expression in French have coded for the presence or absence of a LTI in future-time contexts. For this particular variable, coding practices tend to be underspecified, and although most examples given show the temporal adverbial in the same clause as the future-time context, it is not clear if all examples of adverbial specification are clause immediate. As discussed previously, as well as by Grimm (2016), there is good reason to expand this definition to include LTIs beyond the immediate clause: Temporal reference established by a LTI can be maintained across long stretches of discourse. Thus, in our second reanalysis, we will apply two different operationalizations of what is meant by the presence versus the absence of a LTI in future-time contexts to a dataset of casual conversations. The first coding schema (a clause-level definition) was adopted in Edmonds et al. (2017), whereas the second (a discourse-level definition) is presented here for the first time.

#### *Presentation of original study*

In Edmonds et al. (2017), we examined how NSs and NNSs of French expressed future-time reference in informal oral conversations. Ten English NSs with high L2 (near-native) proficiency were recruited. All had been living in France for at least 4 years. Each participant selected a NS of French with whom they were comfortable speaking in French (e.g., spouse, friend, co-worker). Every dyad was instructed that they would be left alone for 45-60 minutes with a microphone and that they could talk about any subject that they wished. The resulting corpus contains over 8 hours of recordings (more than 77,000 words), with the

NSs and NNSs contributing similar numbers of words (see Donaldson, 2012, for additional details on this corpus). A concept-oriented approach was adopted for the coding of each transcript (Bardovi-Harlig, 2007; von Stutterheim & Klein, 1987). Such approaches “are interested in the range of linguistic devices that speakers use to express a particular concept” (Bardovi-Harlig, 2007, p. 58). In the case of Edmonds et al. (2017), the analysis aimed to identify how NSs and NNSs expressed the concept of future time. In order to do so, we began by identifying all future-time contexts, which was defined as any finite predicate set to occur after the moment of speaking. Each transcript was coded minimally by two researchers and any disagreements were resolved with the third researcher. A total of 947 future-time contexts were identified in the ten transcripts.

All future-time contexts were coded for a variety of variables, including the presence of a specific or non-specific LTI. In Edmonds et al. (2017), the temporal adverbial had to occur within the same clause as the future-time context in order to be coded as present. Following this coding schema, we found that 24.2 percent of future-time contexts in the NS portion of the corpus were accompanied by a LTI. The percentage for the NNSs was similar (23.2%). We then looked more specifically at the co-occurrence of futurate present forms with temporal adverbials. We found that 40.8 percent of futurate present forms produced by NSs and 32.2 percent of such forms produced by NNSs were used in the presence of a clause-immediate temporal adverbial. Thus, we found that use of LTIs was higher with futurate present forms than with the full corpus for both groups of speakers. However, the rates of co-occurrence between futurate present forms and temporal adverbials were strikingly lower than those reported by researchers such as Roberts (2012), who found near categorical co-occurrence of LTIs with futurate present forms in a corpus of native Hexagonal French.

*Reanalysis*

We sought to compare the coding of LTI at the clause level adopted in Edmonds et al. (2017) with a new coding of this variable, which takes discourse context as its scope. In examining the transcripts of the ten conversations, it was clear that future temporal reference was often established by one speaker and maintained by both as they continued to speak about the same topic. One such example is provided in (5), with all verbs occurring in future-time contexts indicated in bold. In this example, the NS speaker has introduced a new topic, namely a play that will be performed on the 28<sup>th</sup> by middle school students. The LTI – *le vingt-huit* ‘the 28<sup>th</sup>’ – appears in the NSs’ first turn, but the talk that occurs in the seven turns that follow maintains this temporal reference.

(5) NS1: *et le vingt-huit **il y a** une avant-première **y a** une pièce de théâtre du collège Jeanne d’Albret*

NNS1: *ah ouais **j’y vais** et Chloé **elle joue**?*

NS1: *ouais **Chloé y joue** ouais*

NNS1: *et **c’est** avant quel film alors?*

NS1: *avant Kelly euh j’sais pas*

NNS1: *ah oui le secret de Kelly ou je sais pas*

NS1: *oui je sais pas ce que c’est*

NNS1: *oui une petite fi oui oui d’accord et **Chloé va jouer**? [oui] **tu vas y aller alors***

NS1: *and the 28<sup>th</sup> **there is** premiere **there is** a play by the Jeanne d’Albret middle school*

NNS1: *ah yeah **I’m going** and is **Chloé performing**?*

NS1: *yeah **Chloé’s performing** yeah*

NNS1: *and so **it’s** before which movie?*

NS1: *before *Opal* uh I dunno*

NNS1: *ah yes *Opal Dream* or I don’t know*

NS1: yes I don't know what it is

NNS1: yes a little gir yes yes ok and **Chloé is going to perform?** [yes] **so you're going to go**

In applying our original coding schema for LTI, only the very first future-time context (*il y a une avant-première*) was coded as being accompanied by the temporal adverbial *le vingt-huit*. For the reanalysis of this independent variable, we redefined what we understood by “presence” of a LTI, extending its scope from the clause level to the discourse context. For this reanalysis, a future-time context is considered to be accompanied by a LTI when, in the same discourse context, a temporal adverbial is used that establishes when that future event is expected to occur. Discourse context was defined as a theme that extends more than one clause and provides continuity to a part of the conversation (cf. Givón, 1983; Lambrecht, 1994; Reinhart, 1981). We began by identifying discourse contexts across the ten transcripts. We then reexamined each future-time context identified in our initial analysis, searching for LTIs within the discourse context. For this coding, we followed previous research and distinguished between specific and non-specific adverbials. Logically, all LTIs identified using our original clause-immediate coding were also included in this new operationalization of the variable. If we return to the example provided in (5), all eight future-time contexts were coded as occurring in the presence of a LTI for the reanalysis, as *le vingt-huit* is when each is expected to occur. Table 4 presents the results from the clause-level coding of LTI (Edmonds et al., 2017) and the discourse level coding (reanalysis).

Table 4. *Distribution of LTI in Edmonds et al. (2017) and in the reanalysis.*

Group	Edmonds et al. (2017): Clause level						Reanalysis: Discourse level					
	Specific		Nonspecific		Absent		Specific		Nonspecific		Absent	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
NSs	81	18.1	27	6	339	75.8	132	29.5	32	7.2	283	63.3

NNSs	82	16.4	34	6.8	384	76.8	149	29.8	50	10	301	60.2
------	----	------	----	-----	-----	------	-----	------	----	----	-----	------

---

The change in scope for the variable of LTI unsurprisingly led to the identification of more instances of the presence of LTI. Whereas we reported that 24.2 percent of future-time contexts in the NS corpus and 23.2 percent in the NNS corpus were accompanied, within the same clause, by a LTI, these percentages increase to 36.7 for NSs and 39.8 for the NNSs in the reanalysis. For NSs, 56 predicates were reanalyzed as occurring within the discourse frame established by the time adverbial, and 83 in the case of NNSs. When we narrow in on uses of the futurate present only, we also see a clear increase in the co-occurrence of this form with a LTI. Whereas Edmonds et al. (2017) reported that futurate present forms produced by NSs were accompanied 40.8 percent of the time by a clause-immediate LTI and that the figure for NNSs was 32.2 percent, a discourse-level definition of LTI brings these percentages to over 50 percent: 52.5 percent for NSs (74 out of 141 instances of futurate present forms) and 52.7 percent for NNSs (77 of 146 total instances of present-for-future).

A comparison of the results from the two approaches to operationalizing the presence of LTI brings to light two observations. First, as concerns questions of SLA, we note that the NNSs who participated in these conversations show similar results to those found for the NSs, regardless of how adverbial specification is operationalized. In other words, in terms of both clause-level and discourse-level patterns, this group of NNSs closely approximates the distributional patterns seen in the NS corpus. This is already remarkable, as one of the challenges inherent in the acquisition of variable structures in a L2 is the acquisition of the rates of use of each variant (Gudmestad, 2014). For the second observation, we consider our results within the larger literature on future-time reference in French. We note that although the new definition of the LTI variable increases the number of present-for-future tokens that occur in the company of such adverbials, the rates reported remain far from an “almost categorical cooccurrence with future adverbials”, which is how Roberts (2012, p. 97)

described the use of the futurate present in his corpus of Hexagonal French. Thus, whether defined as occurring within the clause or within the discourse context, the results from our analysis are clearly different from those reported by Roberts (2012) for Hexagonal French. We presume that this difference either reflects actual differences between the two datasets (in data type or in participant characteristics) or is an artifact of diversity in data coding of the dependent variable in the two studies. As concerns the first possibility, it is plausible that the interview data analyzed by Roberts (and in most other variationist analysis mentioned in this chapter) may show distinct ways of establishing and maintaining temporal reference, likely dependent on the interviewer, when compared to informal conversation between peers who know each other well. This hypothesis suggests that genre or task type may be a factor that should be taken into consideration in future research (Geeslin, 2010; Schilling, 2013). It may, however, also be the case that the differences in rate of adverbial specification derive at least in part from the coding of the dependent variable (i.e., what counts as a future-time context). Indeed, little detail is provided in Roberts (2012) that would allow us to determine how present-for-future forms were identified. In other words, it is impossible for us to know if the two studies went about identifying future-time contexts in the same way (see Polio & Gass, 1997, p. 502, for a similar point concerning replication studies). If the identification criteria for the dependent variable adopted in Roberts (2012) and in Edmonds et al. (2017) differ, this may help explain the divergence in our conclusions.

### **Reflections on data coding**

Decisions made when researchers code their data form a prism through which interlanguage is analyzed and interpreted. One reflection of the importance of these decisions is the space dedicated to data coding in how-to guides to conducting SLA research (Dörnyei, 2007; Mackey & Gass, 2005). However, publications adopting a more critical posture

concerning data-coding practices are less common (cf. Mackey 2012; Myles & Cordier, 2017; Révész, 2012), despite the fact that all researchers are confronted with these difficult questions. In the current chapter, we have attempted to provide two compelling examples from the literature on future-time expression in L2 French demonstrating the room for improvement that exists with respect to consistency and transparency in data coding. In the first example, we showed that the coding of temporal distance has been subject to many different approaches and that this variety in coding practices has two important results. First, it is difficult to build upon previous research, as the categories of this variable are not always the same. Second, we showed, using data published in Edmonds and Gudmestad (2015), that the way in which temporal distance is coded leads to different – and in this case complementary – conclusions about the development of future-time expression in L2 French. We argued that conclusions drawn on the basis of the two approaches to coding temporal distance were valid and of interest. In the second example, we returned to conversation data that we had analyzed for Edmonds et al. (2017). The coding of spontaneous production data comes with many challenges, and in the current chapter, we revisited decisions made concerning the LTI variable. A discourse-level definition of LTI allowed us to better account for how future-time reference was established and maintained across discourse in our dataset. Moreover, results from both the clause-level and the discourse-level analyses of LTI showed that NNSs used temporal adverbials at rates similar to those found for the NSs. However, when we compared the percentage of futurate present forms accompanied by a LTI for NSs and NNSs with what has been reported elsewhere for Hexagonal French (Roberts, 2012), we saw that even with the new discourse-level operationalization of the LTI variable, our results are very different. At this point, it is impossible to know whether the differences reported reflect actual differences in the two datasets and, thus, warrant additional research, or whether these differences are artifacts of coding decisions that were made (with respect to the

dependent variable, in this case). This observation is crucial to our argument. Unless coding practices are clear and decisions justified, readers and researchers find themselves in the position of having to take the authors at their word. Without sufficient detail, which may entail including fully coded contextualized examples, sample items, and/or full tasks, we cannot verify coding decisions, nor can we be sure to be able to replicate what has been done in previous research (as was the case with respect to the LTI variable). And although departures from previous coding decisions may be justified, it is too often the case that such decisions are left unexplained (as was the case for temporal distance, including in some of our own previous research). To this end, we hope that this chapter has succeeded in illustrating the importance of data-coding decisions and practices within SLA.

## References

- Bardovi-Harlig, K. (2007). One functional approach to second language acquisition: The concept-oriented approach. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 57-75). London: Routledge.
- Blondeau, H. (2006). La trajectoire de l'emploi du futur chez une cohorte de Montréalais francophones entre 1971 et 1995. *Revue canadienne des langues vivantes*, 9, 73-98.
- Blondeau, H., Dion, N., & Ziliak Michel, Z. (2014). Future temporal reference in the bilingual repertoire of Anglo-Montrealers: A twin variable. *International Journal of Bilingualism*, 18(6), 674-692.
- Blondeau, H., & Labeau, E. (2016). La référence temporelle au futur dans les bulletins météo en France et au Québec : regard variationniste sur l'oral préparé. *Canadian Journal of Linguistics*, 61(3), 240-258.

- Comeau, P. (2011). *A window on the past, a move toward the future: sociolinguistic and formal perspectives on variation in Acadian French*. Unpublished doctoral dissertation, York University.
- Comeau, P., & Villeneuve, A.-J. (2016). Future temporal reference in French: An introduction. *Canadian Journal of Linguistics*, 61(3), 231-239.
- Corder, S. P. (1981). *Error analysis and interlanguage*. Oxford: Oxford University Press.
- Donaldson, B. (2012). Syntax and discourse in near-native French: Clefts and focus. *Language Learning*, 62, 902-930.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.
- Edmonds, A., & Gudmestad, A. (2015). What the present can tell us about the future: A variationist analysis of future-time expression in native and nonnative French. *Language, Interaction, Acquisition*, 6(1), 15-41.
- Edmonds, A., Gudmestad, A., & Donaldson, B. (2017). A concept-oriented analysis of future-time reference in native and near-native Hexagonal French. *Journal of French Language Studies*, 27(3), 381-404.
- Geeslin, K. L. (2010). Beyond 'naturalistic': On the role of task characteristics and the importance of multiple elicitation methods. *Studies in Hispanic and Lusophone Linguistics*, 3, 501-520.
- Geeslin, K. L., & Long, A. (2014). *Sociolinguistics and second language acquisition: Learning to use language in context*. London: Routledge.
- Givón, T. (1983). *Topic continuity in discourse: A quantitative cross language study*. Amsterdam: John Benjamins.
- Grimm, D. R. (2010). A real-time study of future temporal reference in spoken Ontarian French. *University of Pennsylvania Working Papers in Linguistics*, 16(2), 83-92.

- Grimm, D. R. (2015). *Grammatical variation and change in spoken Ontario French: The subjunctive mood and the expression of future temporal reference*. Unpublished doctoral dissertation, York University.
- Grimm, D. R. (2016). Le présent à valeur de futur en français parlé. *Actes du congrès annuel de l'Association canadienne de linguistique*,
- Grimm, D. R., & Nadasdi, T. (2011). The future of Ontario French. *Journal of French language studies*, 21, 173-189.
- Gudmestad, A. (2014). Variationist approaches to second language Spanish. In K. L. Geeslin (Ed.), *The handbook of Spanish second language acquisition* (pp. 80-95). Malden, MA: Wiley-Blackwell.
- Gudmestad, A., & Edmonds, A. (2016). Variable future-time reference in French: A comparison of learners in a study-abroad and a foreign-language context. *Canadian Journal of Linguistics*, 61(3), 259-285.
- Gudmestad, A., Edmonds, A., Donaldson, B., & Carmichael, K. (in press). On the role of the present indicative in variable future-time reference in Hexagonal French. *Canadian Journal of Linguistics*.
- Gudmestad, A., & Geeslin, K. L. (2013). Second-language development of variable future-time expression in Spanish. In A. Carvalho & S. Beaudrie (Eds.), *Selected Proceedings of the 6th Workshop on Spanish Sociolinguistics* (pp. 63-75). Somerville, MA: Cascadilla Proceedings Project.
- King, R., & Nadasdi, T. (2003). Back to the future in Acadian French. *Journal of French language studies*, 13, 323-337.
- Labov, W. (1966). *The social stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.

- Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Amsterdam: John Benjamins.
- Le Goffic, P., & Lab, F. (2001). Le présent « pro futuro ». *Cahiers Chronos*, 7, 77-98.
- Mackey, A. (2012). Why (or why not), when, and how to replicate research. In G. Porte (Ed.), *Replication research in applied linguistics* (pp. 21-46). Cambridge: Cambridge University Press.
- Mackey, A., & Gass, S. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Erlbaum.
- Marsden, E., Mackey A., & Plonsky, L. (2016). The IRIS Repository: Advancing research practice and methodology. In A. Mackey & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS Repository of Instruments for Research into Second Languages* (pp. 1-21). New York: Routledge.
- Moses, J. G. (2002). *The development of future expression in English-speaking learners of French*. Unpublished doctoral dissertation, Indiana University.
- Myles, F., & Cordier, C. (2017). Formulaic sequence (FS) cannot be an umbrella term in SLA. *Studies in Second Language Acquisition*, 39, 3-28.
- Nadasdi, T., Mougeon, R., & Rehner, K. (2003). Emploi du 'futur' dans le français parlé des élèves d'immersion française. *Journal of French Language Studies*, 13, 195-219.
- Polio, C., & Gass, S. M. (1997). Replication and reporting: A commentary. *Studies in Second Language Acquisition*, 19, 499-508.
- Poplack, S., & Dion, N. (2009). Prescription vs. praxis: The evolution of future temporal reference in French. *Language*, 85, 557-587.
- Poplack, S., & Turpin, D. (1999). Does the *futur* have a future in (Canadian) French? *Probus*, 11, 133-164.

- Ranta, L., & Lyster, R. (2007). A cognitive approach to improving immersion students' oral language abilities: The Awareness-Practice-Feedback sequence. In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 141-160). Cambridge: Cambridge University Press.
- Reinhart, T. (1981). Pragmatics and linguistics: An analysis of sentence topics. *Philosophica*, 27, 53–94.
- Révész, A. (2012). Coding second language data validly and reliably. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 203-221). Chichester, UK: Wiley-Blackwell.
- Roberts, N. S. (2012). Future temporal reference in Hexagonal French. *University of Pennsylvania Working Papers in Linguistics*, 18(2), 97-106.
- Roberts, N. S. (2016). The future of Martinique French: The role of random effects on the variable expression of futurity. *Canadian Journal of Linguistics*, 61(3), 285-313.
- Schilling, N. (2013). Investigating stylistic variation. In J. K. Chambers & N. Schilling (Eds.), *The Handbook of language variation and change* (pp. 327-349). Chichester, UK: Wiley-Blackwell.
- Tagliamonte, S. A. (2012). *Variationist sociolinguistics: Change, observation, interpretation*. Malden, MA: Wiley-Blackwell.
- Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research: “Clozing” the gap. *Studies in Second Language Acquisition*, 33, 339-372.
- Villeneuve, A.-J., & Comeau, P. (2016). Breaking down temporal distance in a Continental French variety: Future temporal reference in Vimeu. *Canadian Journal of Linguistics*, 61(3), 314-336.

von Stutterheim, C., & Klein, W. (1987). A concept-oriented approach to second language studies. In C. W. Pfaff (Ed.), *First and second language acquisition processes* (pp. 191-205). Cambridge, MA: Newbury House.